情報爆発に対応する高度にスケーラブルなモニタリングアーキテクチャ

研究代表者	中島　達夫	早稲田大学・理工学術院・教授
研究分担者	村岡　洋一	早稲田大学・理工学術院・教授
	後藤　滋樹	早稲田大学・理工学術院・教授
	山名　早人	早稲田大学・理工学術院・教授
	甲藤　二郎	早稲田大学・理工学術院・教授

１．研究概要

大規模な分散システムを安定して動作させるためにはシステムが置かれた状況を理解することを可能とする必要がある．本研究では，そのためのインフラストラクチャとして様々な実時間に生成された情報を収集，分析することを可能とするためのモニタリングアーキテクチャに関する研究をおこなう．

２．モニタリングアーキテクチャ

分散アプリケーションは，ネットワーク上の複数のコンピュータにまたがって動作するため，性能解析が難しいという問題がある．単一のコンピュータで動作するイベントトレーサや，特定の環境，特定の分散アプリケーションに特化させたイベントトレーサは存在するが，汎用的に利用できる分散アプリケーション用のイベントトレーサはない．本研究では，各コンピュータで動作するイベントトレーサで採取したログを集約することで，分散アプリケーションの性能解析をする手法を提案する．

単一コンピュータ用のイベントトレーサを用いて分散アプリケーションの性能解析をする場合，それぞれのコンピュータ毎に別々に記録されたログを解析する必要がある．そのため，異なるコンピュータで動作するアプリケーション間でのデータの送受信など，アプリケーション同士の関係を把握することが難しい．Pip などの分散環境用の性能解析ツールも存在するが，事前に解析対象である分散アプリケーションの動作を定義する必要があったり、ネットワークの障害発見に特化していたりするなど、汎用的な分散アプリケーション用の性能解析ツールはない．

本システムでは、分散アプリケーションが動作しているそれぞれのコンピュータで，単一コンピュータ用のイベントトレーサを用いてログを採取し，それらのログを集約することで，アプリケーション間の関係も含めて性能解析を行う。

各ログの順序関係を維持することは，重要である．なぜなら，アプリケーションで問題が発生した場合，その原因となる事象は，問題よりも前に発生しているからである．ログから問題の原因を探すには，問題が発生した時点から遡って解析していく．そのため，問題よりも，原因が後に起こったこととして記録されていた場合，原因を発見することはできない．しかし、複数のコンピュータで採取したログを集約する場合，ログのタイムスタンプによっ

て並び替えるだけでは，ログの順序関係を維持することはできない．なぜなら，それぞれのコンピュータが自立した時計を持っているため，ログのタイムスタンプがずれ，順序が入れ替わる可能性があるからである．

複数のログの順序関係を維持するには，コンピュータ間のパケット通信のログを用いる．コンピュータ A のログに，パケット p の送信ログ s，コンピュータ B のログに，p の受信ログ r がある場合，r よりも s の方が先に発生したことがわかる．ただし，送信ログ s で送信したパケットと，受信ログ r で受信したパケットが同一であることがわかる情報をログに残す必要がある．

パケット通信のログを調べ，パケット送信ログのタイムスタンプよりも，パケット受信ログのタイムスタンプの時間が早ければ，順序関係を正すためにタイムスタンプを調節する必要がある．タイムスタンプを調節する手法として，いくつかのアルゴリズムが考えられる．ひとつは，ランポートのアルゴリズムを用いたものである．また，別の方法として，NTP で用いられている方法もある．これは，パケットのラウンドトリップ時間の統計をとることで，二つのコンピュータ間の通信遅延時間を推定し，それを元に時間を修正する，という方法である．今回，Linux 用のイベントトレーサである Linux Trace Toolkit Next Generation (LTTng)と，ログ表示解析ツールである Linux Trace Toolkit Viewer (LTTV)をベースとして，分散環境における性能解析システムを実装した．まず，ログの対応関係をとるため，パケット送受信時に，IP，TCP，UDP の各ヘッダとソケットの情報を記録するシステムを LTTng 上に実装した．また，送受信元 IP アドレス，ポート番号，シーケンス番号など，記録した情報からパケット送受信ログの対応関係を求め，ランポートのアルゴリズムを用いてタイムスタンプの調整を行う．さらに，各コンピュータのプロセス間でやりとりされるパケットを，プロセスの状態と併せて表示する，LTTV 用のモジュールを作成した．これによって，複数のログを一つの画面にまとめて表示することができる．

また，本研究では，LLTng を利用して実時間システムの性能上の問題を発見する実験をおこなった．実時間システムでは，正確に決められた時間にタイマーをトリガーすることが重要である．しかし，Linux のような複雑なＯＳでは様々な要因によりタイマーのトリガーの遅延が生じる．現状では，その詳細な原因を追究した研究は存在しないので，Linux の実時間システムへの適用可能性は不明確なことが多かった．しかし，本研究では，LTTng を利用することにより，タイマーのトリガーに遅延が生じる場合の原因を明確にした．

３．ローカルノードに関する詳細な情報取得を可能とするマイクロカーネル

本研究では，ローカルノードに関する詳細な情報取得を可能とするマイクロカーネルとして，分散処理，ネットワーク処理に適した機能を持つ汎用 OS と実時間処理に適した機能を持つ実時間 OS（RTOS）の共存を可能にするマイクロカーネルの研究を行っている．

ローカルノードに関する詳細な情報の取得には実時間処理が必要とされ，そのための機能を提供する RTOS が使用できることが望ましい．例えば，RTOS ではタスクの優先度に従っ

たスケジューリングが行われるため，重要な情報収集がそれほど重要でない情報の収集に阻害される場合を出来る限り排除することができる．また，定期的に情報を収集するためには，RTOSの提供する周期タスクの機能を用いることで，精確な時間に情報の収集が可能になる．この時間の正確さも優先度に従った順になり，重要な情報の収集のタスクはより精確な時間に実行可能になる．

一方，取得したローカルノードに関する情報を他ノードやサーバ等に提供するため，また情報をもとに協調した処理を行うためには，分散処理，ネットワーク処理に適した機能を持つ汎用 OS が必要となる．RTOS もネットワーク処理が可能なものもあるが，多くの場合その機能は制限されたものであり，広域環境における相互接続性，分散処理を行ううえで必要となるミドルウェアの機能といった面から，汎用 OS の使用が望ましい．

そこで本研究では汎用 OS と RTOS の共存を可能にするアーキテクチャの研究を行っている．具体的な実装としては，汎用 OS として Linux、RTOS としてμITRON を用い，これらを仮想マシンモニタ（VMM）上に実装するハイブリッド環境を構築している．また，VMM を軽量化し、また VMM 上で OS を動作させるにあたって OS に適した仮想化を行なうことで，実装および実行オーバヘッドを抑える．

Linuxは大きく複雑なOSであり，またバグフィックスや機能拡張のためにバージョンアップが頻繁である．従って，paravirtualizationを用いた場合，導入コスト，維持コスト共に非常に大きくなる．そのため，できるだけ単純な改変で VMM 上に汎用 OS を動作させることができる仮想化手法を開発した．一方，μITRON は Linux と比較するとはるかに小さく単純なである．また，通常 RTOS には安定性が最優先されるためバージョンアップは希である．また，小さな組み込み用プロセッサでも動作するようになっているため，プロセッサの複雑な機能を使用しない．従って，paravirtualizationを導入したとしても，改変量はわずかで済む．

４．分散システム情報収集分析ミドルウェアの研究

本年度は，昨年度からの自己組織型セキュリティミドルウェアに関する検討をさらに進め，ネットワーク構成技術の面からは，ノード安定度を導入した自律分散的な経路制御プロトコルの提案と評価を行い，提案によるデータ配信の安定性向上効果を示した．また，コンテンツ配信の性能，ならびに信頼性向上に帰着する動的コンテンツの管理手段，並びに DHT 網におけるコンテンツの複製手段についても検討，報告を行った．また，センサーネットワークによる位置推定システム，ならびに混雑推定システムの検討，実験を行い，それぞれについて，無線特性が与える推定精度への影響と安定化手段を報告した．

またネットワーク測定技術の面から，多地点における Dark IP による測定と分析の研究を進めた．さらに物理的なセンサーマシンを不要とする仮想センサーの技術を確立して，実センサーと同様の分析を行えることを実証した．また，ネットワークの通信品質(QoS)に関しては，多数のマシンが同時に稼働する場合の品質評価をシミュレーションにより行った．ネ

ットワークのアプリケーションの観点からの測定技術の実証を進めて，ユーザの閲覧行動を分析するための測定法について実証実験を行った．

ノード安定度を導入した自律分散的な経路制御プロトコルの評価実験では，P2P 型（オーバーレイ型）の配信木を構成する各ノードに対して，帯域幅で重み付けた長時間セッション履歴に基づく安定度を定義し，この安定度をメトリックする配信木の構成方式を提案した．従来手法としてセッション時間＊帯域幅やコンテンツのアップロード回数をメトリックとする方式が知られているが，前者は長時間履歴を考慮しておらず，後者は帯域幅を考慮していなかった．また，昨年度は予備経路構築による（短期的な）障害対策を提案，評価しており，長期的な障害対策となる本方式を組み合わせることで，より可用性の高い P2P 型データ配信が可能になる．また，安定度の定義は任意であり，現在は信頼性向上により直接的に貢献するメトリックの検討と，より柔軟なメッシュ網拡張の検討を行っている．

このほか，動的コンテンツの管理手段，並びに DHT 網におけるコンテンツの複製手段についても検討，報告を行った．前者は動的に内容が更新されるコンテンツの複製手段に関するものであり，コンテンツを細分化し，その部分毎の更新先を人気度とネットワーク状況に応じて決定する．後者は DHT に代表される構造化 P2P におけるコンテンツ複製に関するものであり，コンテンツの人気度に応じて検索と複製を協調させることで，より効率的な負荷分散を図るものである．共に，コンテンツの複製手段に関するものであり，ノード障害への耐性も提供する．

また，センサーネットワークに関して，自立走行型ロボットを利用した位置推定アルゴリズムに対する検討，実験を行った．具体的には自身の位置と通信相手の位置を同時に推定する SLAM と呼ばれるアルゴリズムを利用し，ロボットに内蔵されたセンサー（加速度等）と複数の通信相手からの無線強度を組み合わせた位置推定の実験評価を行った．その結果として，通信相手の配置によって精度が変わること（特に人の往来によって変動する，無線特性の安定性が精度に大きく影響すること，などを明らかにした．一方，後者の無線特性の変動を逆に利用し，無線センサーによる居室内の混雑度測定への応用についても検討，実験を行った．人が静止している場合と動いている場合の無線特性の違いも評価を行い，無線特性の変動から，ある程度の混雑度の推定と人間の挙動が把握できることを明らかにした．さらには，Web サービスベースの簡易なアプリケーションの試作も行なった．

ネットワークの測定に関しては、昨年度にはソフトウェアによる測定性能の向上をはかった．本年度は多数地点における測定を実際に行い，データの分析を行った．ここで用いた技法は Dark IP と呼ばれるものである．すなわち，通常のマシンには割り当てられていない IP アドレス，未使用のはずのアドレスを観測する．このようなアドレスには通常のパケットは到着しない筈である．実際に観測をしてみると，ウィルスが散布するパケット，不正侵入の前段階として行われるポートスキャン，ネットワークの設定を誤ったために発生するパケットなどが捕捉される．このようなデータを分析することにより，ネットワークの上で脅威と考えられている不正なパケットを判別することができる．特に多地点において同時に観測す

ることにより，広域ネットワークにおける特異な現象を把握することができる．

さらに本年度は，上記の測定方法を改良するために，測定用の物理的なセンサマシンを設置せずに測定を可能とする方法を考案した．この方法を用いる場合には，パケットが通過するルータの flow data を分析する．flow data は多くのルータが標準的に備えている機能である．本研究の技法の特徴は，flow data の中から Dark IP に相当するデータを抽出するところにある．今回提案した新しい方法を用いて，センサーマシンを設置した場合と同様のデータの収集と分析ができることを示した．

通信品質の測定については，パケットロスや通信遅延が影響を与えやすい音声の通信について，通信品質の測定方法を研究した．昨年度には，ごく小規模な実験用ネットワークを用いて通信品質(Quality of Service，QoS)を保証する技術を研究した．本年度は多数のマシンが同時に通信する場合を検討するために，シミュレーションにより通信品質を測定し，QoS を保証するための設定方法を研究した．

ネットワークのアプリケーションレベルの測定としては，Web の閲覧履歴の特徴を記述する方法を考案した．本研究では，閲覧動作の特徴を時系列で分析する，たとえば，一人の利用者の閲覧履歴でも，通常と異なる動作をした場合には，それを発見することができる．

今後の展望は下記のとおりである．

- 自律分散的な経路制御プロトコルに関しては，国際学会の動向として，対象がより柔軟性の高いメッシュ系トポロジーに移行しており，筆者らもまたメッシュ網拡張に重点を置いた検討を進めていく．コンテンツ複製に関する検討も同様である．一方，センサーネットワークに関しては，現時点では高々数ノードの検討にとどまっており，より大規模な実験，評価を進めていく予定である．
- ネットワークの測定に関しては，本年度に確立した Dark IP と flowdata による仮想センサとの比較を行い，仮想センサによる測定精度を向上する予定である．またアプリケーションレベルの測定については，利用者の閲覧履歴の特徴記述と，コンピュータが自動的に行う通信の履歴が同様の枠組で把握できる筈である．この観点で測定技術として一般化をはかる．

５．自己組織型セキュリティミドルウェア

平成 19 年度は，平成 18 年度に研究開発を行った「情報の生成（あるいは出現）シーケンスに着目した分析手法」を，様々なアプリケーションに対して適用し，その効果を検証した．なお，平成 19 年度はノイズの少ないデータを対象として検証を行った．

まず，テキストを対象とした類似度判定への適用例として，Web 上のデータを対象に著作権違反コンテンツの発見が可能かどうかについて検証を行った．図 1 に全体構成を示すように，seed text（著作権を持つオリジナルテキスト）を入力すると Chunker により文節単位に分解され，次に Query Generator で文節単位の n-gram が出力される．本システムでは n と

して 3 を採用している。Query Generator で生成された 3-gram は商用検索エンジンへ Query として送信され，検索結果の URL に基づき Candidate page Getter により当該 URL の Web ページが取得される．取得されたページを式(1)により類似度判定し類似度が 0.3 以上のものを著作権違反候補ページとした．式(1)における S は seed text であり，C は Candidate page Getter により得られた Web ページである。文節の配置順序，すなわちシーケンスに着目し，seed text と一致する最長のシーケンスを抽出し類似度判定に用いている．このようにシーケンスに着目することで，テキスト中に入ったコメントや改行等を無視した類似度判定が可能となる．

$$Sim(S,C) = \log_2 \left\{ \frac{|Lcs(S,C)|}{|S|} + 1 \right\} \quad \ldots (1)$$

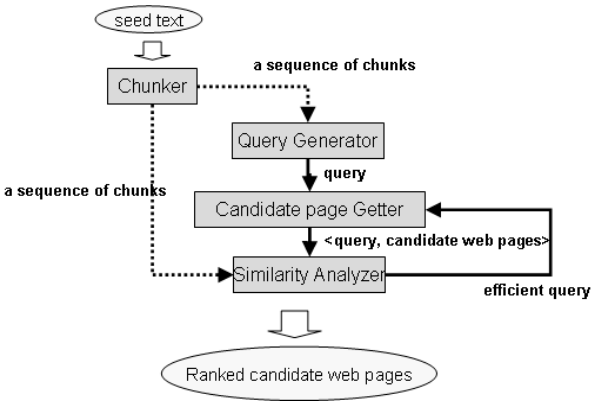

図 1　著作権違反コンテンツ発見システムの全体構成図

結果は表 1 に示すように，精度 94%を達成することができた．これにより，シーケンスに基づく類似度判定が Web ページのようなコメント等が間に多数挿入されるようなテキストデータに対しても有効であることを示した．

表 1　実験結果

	Related Pages			Unrelated Pages	Total
	>80%	30-80%	<30%		
English news	1,658	334	26	28	2,046
English lyrics	1,075	272	16	160	1,523
Japanese news	213	68	19	72	372
Japanese lyrics	1,240	354	36	62	1,692
Total	4,186 (74%)	1,028 (18%)	97 (2%)	322 (6%)	5,633 (100%)
	Precision =94%				

その他，平成 19 年度は, Web 上のプロフィールデータに対する類似度判定への適用，翻訳支援システムにおける冠詞誤り発見への適用，オークション詐欺検知モデルへの適用，商用検索エンジンのランキングの時系列変化解析への適用を行い，提案手法の効果を確認した．

６．センサーを利用した応用サービス

センサー情報を利用するサービスとしてAmbient Lifestyle Feedback System を提案した．本システムは，ユーザの行動をユーザに意識させずにモニタリングして，その行動を改変するためにフィードバック情報をユーザに心理的負荷を課さないように提示する．

ユーザの行動を認識するためには，ユーザが使用する日常物を監視することで，ユーザは特別な装置を身につける必要はない．また，フィードバックの出力は水槽や絵画などの日常の生活の中で違和感が少ないものを選択した．また，フィードバック手法に関しては基礎的な行動心理学の理論を用いてユーザが行動を変化させるようにした．

本年度は，提案したアイデアの有効性を示すため仮想水槽システムとモナリザ本棚システムを構築した．また，実際にユーザスタディをおこなうことにより，提案システムの有効性を示した．

７．研究成果リスト

著書、論文

- T. Ueda, T. Hori, Y. Hirate, H. Yamana, Knowledge EngineeringT. Tashiro, "EPCI: Extracting Potentially Copyright Infringement Texts from the Web," Proc. of 16th International Conference on World Wide Web (WWW2007), pp.1151-1152 (2007.5)
- T.Shimoyama and Y.Muraoka, “Two methods for speeding up similarity measurement for profile data”, Proc. of the 2007 International Conference on Information and Knowledge Engineering (IKE’07) (2007.6)
- Yasuaki Yoshida, Takanori Ueda, Takashi Tashiro, Yu Hirate, Hayato Yamana:"What's going on in search engine rankings?", Proc. of the 2008 IEEE International Symposium on Mining the Asian Web(MAW2008) (2008.3)
- 近藤 秀和, 村岡 洋一, “ウェブブラウザ「Lunascape」”, コンピュータソフトウェア, Vol. 24 (2007), No. 4, pp.139-152 (2007)
- Makoto Iguchi and Shigeki Goto, Anonymous P2P web browse history sharing for web page recommendation, IEICE Transactions on Information and System, Vol. E90-D, No.9, pp1343—1353, September, 2007.
- T.Mori, T.Takine, J.Pan, R.Kawahara, M.Uchida, and S.Goto, Identifying

Heavy-Hitter Flows from Sampled Flow Statistics, IEICE TRANSACTIONS on Communications, Vol. E90-B, No.11, pp3061-3072, November, 2007.

- S.Zhou, J.Katto and Y.Yasuda, Scalable Maintenance for Strong Web Consistency in Dynamic Content Delivery Overlays, IEEE ICC 2007, pp.1728-1733, June, 2007.
- Megumi Ito and Shuichi Oikawa, Mesovirtualization: Lightweight Virtualization Technique for Embedded Systems, Proceedings of the 5th IFIP International Workshop on Software technologies for future Embedded and Ubiquitous Systems (SEUS 2007), Springer-Verlag LNCS 4761, pp496–505, 2007.
- 伊藤愛、追川修一, Gandalf VMM における Shadow Paging の実装と評価, **情報処理学会第 19 回コンピュータシステム・シンポジウム**, 2007.
- Tatsuo Nakajima, Vili Lehdonvirta, Eiji Tokunaga, Hiroaki Kimura. Reflecting Human Behavior to Motivate Desirable Lifestyle. Proceedings of DIS 2008, Cape Town, South Africa (forthcoming).
- Tatsuo Nakajima, Vili Lehdonvirta, Eiji Tokunaga, Masaaki Ayabe, Hiroaki Kimura, Yohei Okuda. Lifestyle Ubiquitous Gaming: Making Daily Lives More Plesurable. Proceedings of RTCSA 2007, Daegu, Korea.

受賞

- 吉田泰明：学生発表奨励賞, データ工学ワークショップ DBWS2007(2007.7)
- 平手勇宇：情報処理学会　平成 18 年度 CS 領域奨励賞を受賞(2007.7)